Civil Engineering and Applied Solutions

Research Article

journal homepage: ceas.journals.umz.ac.ir

CEAS Civil Engineering and Applied Solutions

Prediction of Rice Husk Ash-Based SCC Compressive Strength: Data-Driven Framework

Hamid Farrokh Ghatte [a*], Ali Nazari [b]

[a] *Department of Civil Engineering, University of Antalya Bilim, Antalya, Turkey*
[b] *Department of Civil Engineering, Sharif University of Technology, Tehran, Iran*

A R T I C L E I N F O

Keywords:
Prediction
Self-compacting concrete (SCC)
Gene expression programming (GEP)
Machine learning algorithms

Article history:
Received 07 July 2025
Accepted 13 July 2025
Available online 01 August 2025

A B S T R A C T

The construction and upkeep of concrete structures have posed significant technical and financial challenges over the past decade. In response, self-compacting concrete (SCC) has gained attention due to its superior mechanical performance and reduced environmental footprint. This study investigates the application of gene expression programming (GEP) in developing a predictive model for estimating the compressive strength (CS) of self-compacting concrete incorporating rice husk ash (RHA). To assess the model's reliability, its predictions were benchmarked against those from two well-established machine learning methods: multiple linear regression (MLR) and artificial neural networks (ANN). A total of 651 experimental records related to RHA-based SCC were gathered from trustworthy references. The model's performance was then quantified using key statistical measures, including the correlation coefficient (R), root mean squared error (RMSE), and mean absolute error (MAE). The GEP model outperformed the ANN and MLR approaches, delivering greater accuracy and lower error levels. Additionally, the study introduced a gene expression-based formula derived from the GEP model for estimating compressive strength at different curing ages, achieving a correlation coefficient of 0.49 and error values ranging from 5 to 9 MPa, which highlights its strong predictive ability. This equation provides a practical tool for preliminary mix design and the quick assessment of SCC mixtures. Sensitivity analysis revealed that binder content was the most significant parameter influencing CS prediction.

1. Introduction

Self-compacting concrete (SCC) is regarded as one of the breakthroughs in modern concrete engineering practices. As a relatively recent innovation in construction materials, emerging over the past three decades, SCC was made possible by the development of superplasticizers as a new class of admixtures. SCC is a type of concrete with medium to high strength that can spread solely by the force of gravity, completely occupying the formwork without requiring external vibration [1]. In the early 1980s, the increasing complexity of construction projects in Japan, combined with the growing volume of reinforcement, led to poor compaction and decreased execution quality [2]. Given the rising global population and growing human needs, the generation of industrial and agricultural waste is increasing steadily. The construction industry presents a valuable opportunity to utilize such waste materials. This research investigates the application of rice husk a byproduct of agriculture and its ash, which serves as a pozzolanic substance abundant in reactive silica, for potential use in construction materials. Appropriately processed rice husk ash (RHA) has been shown to enhance the durability of concrete against aggressive environments, reduce reinforcement corrosion, and play a significant role in soil stabilization, cementitious block production, and the fabrication of lightweight insulating concrete and bricks [3, 4].

* Corresponding author.
E-mail addresses: ghatte@itu.edu.tr (H. Farrokh Ghatte).

https://doi.org/10.22080/ceas.2025.29625.1025
ISSN: 3092-7749/© 2025 The Author(s). Published by University of Mazandaran.

How to cite this article: Farrokh Ghatte, H., Nazari, A. Prediction of Rice Husk Ash-Based SCC Compressive Strength: Data-Driven Framework. Civil Engineering and Applied Solutions. 2025; 1(3): 17-28. doi:10.22080/ceas.2025.29625.1025.

In recent years, the emergence of new materials and technologies has introduced a larger number of parameters into concrete mix design, sometimes doubling the number of influencing variables. Consequently, data-driven modeling approaches, particularly those built on experimental data, have attracted significant interest among researchers. Classical modeling methods such as multiple linear regression (MLR) and artificial neural networks (ANNs) have been widely used to solve problems in concrete technology. For instance, Chen and Pai [5] used MLR to predict compressive strength based on physical properties, analyzing various input variable combinations, and reporting results using statistical performance indicators. The capability of ANNs to predict concrete properties has also been assessed in several studies. Sonebi et al. [6] and Kalman Šipoš et al. [7] reported that ANN models could serve as effective alternatives to laboratory testing for estimating both fresh and hardened properties of concrete. Similarly, Wang et al. [8] applied both ANN and MLR methods to model the behavior of concrete and presented their results through statistical analyses.

However, classical data-driven methods often fail to provide explicit predictive equations from the model outputs. Therefore, modern modeling techniques like GEP, which can deliver accurate and interpretable mathematical expressions, have gained attention among researchers [9]. As an illustration, Gholampour et al. [10] utilized datasets related to recycled concrete to construct predictive models using GEP, aiming to estimate key mechanical properties such as compressive strength, tensile strength, and elastic modulus.

In recent years, while many studies have focused on applying machine learning techniques such as multiple linear regression and artificial neural networks for predicting the compressive strength of various types of concrete, limited research has been conducted on the use of GEP for modeling the compressive strength of self-compacting concrete incorporating rice husk ash. Moreover, previous studies have often relied on relatively small datasets or lacked explicit predictive equations that can be directly used for practical mix design. These limitations underscore the need for more comprehensive modeling approaches that provide both high predictive accuracy and interpretable formulations.

The novel contributions of the present study can be summarized as follows:

- Development of a GEP-based predictive model for estimating the compressive strength of SCC containing rice husk ash, trained and validated on an experimental dataset.
- Comprehensive comparison of GEP performance with conventional models (MLR and ANN) to highlight its advantages in terms of accuracy and interpretability.
- Derivation of an explicit mathematical equation from the GEP model to facilitate practical use in preliminary concrete mix design.
- Sensitivity analysis to identify the most influential parameters affecting compressive strength prediction.

2. Methodology

2.1. Multiple linear regression (MLR)

Regression analysis is a statistical approach used to explore how a dependent variable responds to changes in one or more independent variables, while holding others constant. This method provides insights into the nature and strength of these relationships. Typically, regression analysis aims to construct a mathematical model that relates independent variables to the prediction of a dependent variable's value [11, 12]. Today, regression analysis is widely used for predictive modeling. Among various techniques available, linear regression and least squares regression are among the most commonly applied. According to Sobhani et al. [13] linear regression is a form of regression analysis that models the relationship between one or more independent variables and a dependent variable using a linear regression equation. In contrast, the goal of nonlinear regression is to find a suitable nonlinear equation that fits the relationship.

2.2. Artificial neural networks (ANNs)

The artificial neural network (ANN) model known as the perceptron was introduced by Frank Rosenblatt in 1958 [14]. These networks typically consist of three layers: input, hidden, and output. In feedforward neural networks, input signals propagate forward from the input layer through the hidden layers to the output layer [15]. The quantity of neurons in the input and output layers is directly aligned with the number of corresponding input and output variables. There is no definitive rule for determining the number of neurons in the hidden layer; this is generally determined based on problem complexity and through trial and error [16]. To minimize the error, the weights between neurons are updated in a backward direction from the output layer toward the input layer, and this process is repeated iteratively until the desired output is achieved.

2.3. Gene expression programming (GEP)

Gene expression programming (GEP), developed by Ferreira in 1999 [17], is a variant of genetic algorithms that evolves solutions by evaluating individuals through a fitness function and applying genetic operations via one or more operators. While similar in concept to genetic algorithms, which operate on binary strings, GEP differs in that it uses tree-based structures to generate optimal solutions [18]. Ferreira incorporated various operators in GEP, such as mutation and crossover. One of the key features of these operators is their ability to avoid producing invalid individuals by applying error-free operations. GEP utilizes different types of crossover mechanisms, including one-point, two-point, and gene crossover [17]. Among these, the two-point crossover is considered more effective as it enables more frequent activation and deactivation of non-coding regions within chromosomes. GEP represents various phenomena by employing predefined sets of functions and terminal symbols. In this study, the GeneXpro Tools 5.0 software was employed to implement the GEP model. To better illustrate the modeling approach employed in this research, a workflow diagram has been provided in Fig. 1. This flowchart depicts the key steps, including dataset compilation, preprocessing, selection of input parameters, model development using MLR, ANN, and GEP techniques, performance evaluation with statistical metrics, and sensitivity analysis to identify the most influential factors affecting compressive strength prediction.

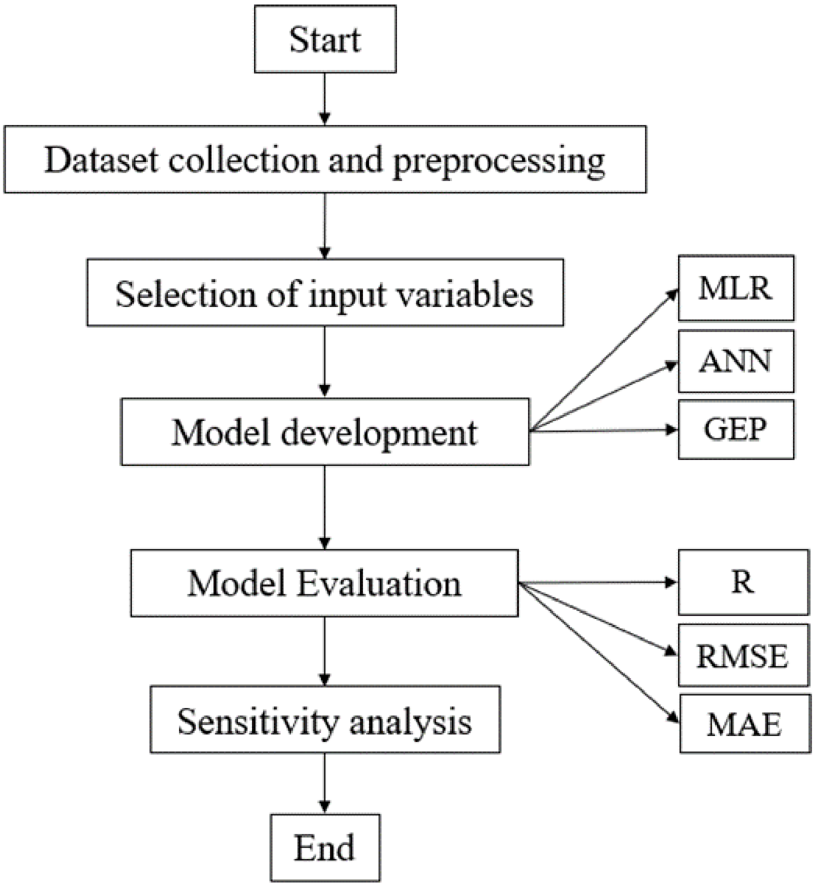


Fig. 1. Workflow of the computational modeling framework adopted in this study.

3. Dataset used in this study

To estimate the CS of SCC incorporating RHA using data-driven models, a comprehensive dataset was collected from various experimental studies. This dataset consists of 156 data points obtained from different laboratory tests related to the CS of SCC incorporating RHA, compiled from previous studies [19-22]. The dataset was split into training and testing subsets, with 75% allocated for model training and the remaining 25% reserved for evaluation. All samples were designed using different combinations of the following materials: fine aggregate (FA), coarse aggregate (CA), cement (C), rice husk ash (RHA), superplasticizer (SP), and water (W). The numerical surveys informed the success in simulating different problems using an ANN [23, 24]. On the other hand, the structural behavior of RC elements and beam-column joints, as well as retrofitting approaches by employing SCC and/or normal concrete, have been additionally investigated through cyclic or monotonic loadings [25-27].

To account for differences in mix proportions and testing conditions, the dataset includes concrete compressive strength values measured at 7, 28, and 90 days under standard curing conditions. The measured CS (MPa) served as the output variable (target), while the seven aforementioned components acted as the input variables. As this dataset was compiled from a variety of published research, it represents a valuable and credible source for modeling purposes. To assess the consistency and distribution characteristics of the data, statistical evaluations were carried out. Table 1 reports the range, mean, standard deviation, and coefficient of variation for each variable. Table 2 outlines the error metrics used to assess the models' predictive accuracy.

Table 1. Summary of descriptive statistics for the considered variables.

Variables	C	W	FA	CA	RHA	AS	SP	CS
Max	560	341	814	1319	135	90	7.33	106.5
Min	0	119	539	995	0	1	0	16.5
Mean	441.54	162.1	608.3	1250	53.1	59.19	4.68	28.92
std dev	58.9	13.95	100.28	92.8	37.1	23.8	93.2	33.89

Table 2. Applied performance evaluation metrics to the models.

No.	Equation
(1)	$R = \frac{\sum_{i=1}^{M}(O_i-\overline{O}).(P_i-\overline{P})}{\sqrt{\sum_{i=1}^{M}(O_i-\overline{O})^2\sum_{i=1}^{M}(P_i-\overline{P})^2}}$
(2)	$RMSE = \frac{\sum_{i=1}^{M}(P_i-O_i)^2}{M}$
(3)	$MAE = \frac{\sum_{i=1}^{M}\lvert P_i-O_i\rvert}{M}$

4. Comparison and evaluation of proposed models

To compare the performance of the proposed models during the training and testing phases, statistical indices including the correlation coefficient (R), root mean square error (RMSE), and mean absolute error (MAE) were used, as presented in Table 2. In these equations, Oi represents the observed values, Pi denotes the predicted values, $\overline{O}$ is the mean of observed values, $\overline{P}$ is the mean of predicted values, and M is the total number of data points.

Additionally, the developed models were compared and evaluated against other data-driven models reported by previous researchers. Specifically, the compressive strength prediction models of Belalia Douma et al. [28], Kaveh et al. [29], Inqiad et al. [30], and Siddique et al. [31] were examined, and their performance was compared with the GEP-based model proposed in this study.

5. Results and discussion

5.1. Determining the optimal input combination for developing predictive models

In this study, an effort was made to identify the best combination of input parameters for developing predictive models. Various combinations of input parameters were tested, and the proposed models were developed accordingly. It should be noted that the parameters were assessed both individually and in combination.

To determine the optimal combination, previous literature and experimental data were reviewed, and different groupings of parameters were analyzed. Based on model performance results, the following input configuration was selected as the best-performing scenario:

1. Compressive strength as the output, and the following as inputs:

Water (W), Cement (C), Fine aggregate (FA), Coarse aggregate (CA), Rice husk ash (RHA), and Superplasticizer (SP).

2. W/C, FA/C, CA/C, SP, and RHA content.

3. W/FA, FA/C, CA/C, SP, and RHA content.

4. W/CA, FA/C, CA/C, SP, and RHA content.

In this study, MLR was used to determine the best combination of input variables for predicting compressive strength. The goal was to select the optimal input set for model development using statistical regression techniques. MLR is known for its simplicity and interpretability and is widely used in preliminary modeling when dealing with real datasets.

The very low error associated with this method suggests that MLR can be reliably used in early-stage model selection before adopting more complex nonlinear algorithms. Among the four tested combinations (shown in Table 3), the fourth input set consisting of W/CA, FA/C, CA/C, SP, and RHA content resulted in the best prediction accuracy.

Table 3. Assessment of suggested scenarios for identifying the optimal set of input variables.

Imput	R	RMSE	MAE
Scenario -1	0.92	6.41	7.71
Scenario -2	0.9	7.49	8.1
Scenario -3	0.83	10.2	11.29
Scenario -4	0.86	9.31	9.6

This configuration achieved an R of 0.92 and the lowest RMSE of 6.41 MPa in predicting compressive strength, indicating its superior predictive performance. Furthermore, when this optimal input set was tested using the ANN and GEP models, the results confirmed its higher reliability compared to the other three input combinations, as illustrated in Fig. 2.

5.2. ANN model for predicting concrete strength

After identifying the best combination of input variables for estimating the CS of SCC with RHA, the ANN modeling process was initiated. A multilayer feedforward ANN architecture was utilized in this study. The quantity of neurons in the hidden layer significantly influences the model's overall performance. To identify the most suitable number of neurons, different values were tested in a trial-and-error process. In this study, a single hidden-layer architecture was used, as prior research has shown it to be sufficient for capturing complex nonlinear relationships [11-13].

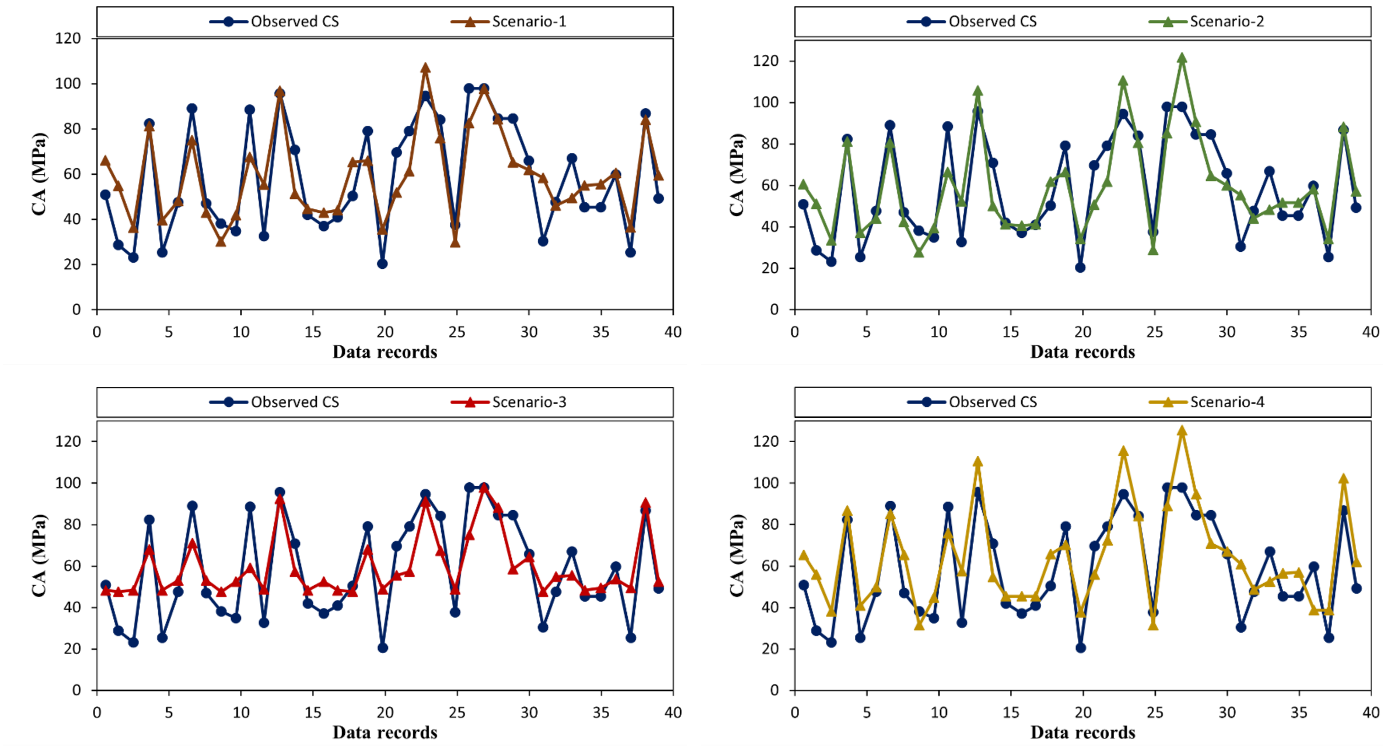


Fig. 2. Time series diagrams corresponding to the suggested alternatives.

The Levenberg–Marquardt (trainlm) algorithm was used for training, given its high efficiency and minimal prediction error. Additionally, the sigmoid transfer function was employed in both the hidden and output layers, which has proven effective in previous studies. For training and testing the networks, 90% (141 samples) and 20% (31 samples) of the data were used, respectively, and the performance of each phase is reported in Table 4. The performance of the ANN model during both phases was evaluated using the same statistical indicators (R, RMSE, and MAE). Fig. 3 illustrates the evaluation metrics for each artificial neural network model in a graphical format. Among the different configurations tested, the optimal ANN model had 5 neurons in the hidden layer, a learning rate of 0.01, a momentum coefficient of 0.9, and was trained for 2000 epochs. It is important to mention that increasing the number of neurons in the hidden layer beyond a certain limit led to overfitting and reduced generalization capability. The best-performing ANN model, as noted, consisted of a single hidden layer with 5 neurons and 2000 training iterations. In this study, ten different ANN configurations (ANN 1 to ANN 10) were developed by systematically varying the number of neurons in the hidden layer from 1 to 10. Each ANN model in Table 4 corresponds to a specific configuration with a distinct number of hidden neurons. This approach allowed us to evaluate the impact of the network complexity on prediction accuracy and to identify the optimal architecture providing the best performance metrics.

Table 4. Performance metrics of ANN models during training and testing phases.

Models	Training			Testing		
	R	RMSE	MAE	R	RMSE	MAE
ANN 1	0.69	15.6	16.6	0.6	15.6	16.6
ANN 2	0.73	11.2	12.8	0.79	13.3	14
ANN 3	0.94	5.3	5.9	0.92	7.5	8.3
ANN 4	0.95	4.5	5.1	0.93	5.1	5.6
ANN 5	0.92	5.4	5.8	0.92	6.4	6.9
ANN 6	0.93	5.8	6.7	0.92	6.9	7.5
ANN 7	0.94	4.9	5.4	0.9	7.1	7.9
ANN 8	0.92	6.46	7.5	0.85	9.9	10.2
ANN 9	0.93	6.34	7.6	0.86	8.5	9.5
ANN 10	0.86	9.76	10.9	0.83	1.9	11.5

5.3 GEP model for predicting concrete strength

In this study, GEP was used to predict the CS of self-compacting concrete containing RHA. The model was developed based on a set of initial input parameters, from which the algorithm started its search, and ultimately generated an explicit mathematical expression for the output. As presented in Table 5, the values assigned to each parameter in the GEP model are shown. This provides a straightforward and interpretable expression that connects mix design parameters with CS.

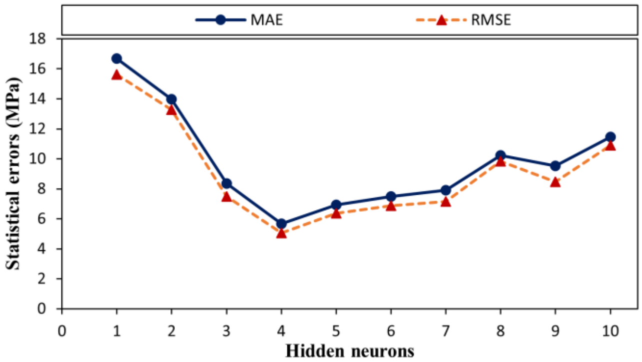


Fig. 3. Assessment of hidden layer neuron counts in ANN configurations.

The final equation, derived using the GEP approach, provides a reliable and practical tool for predicting the compressive strength of SCC incorporating rice husk ash. For the prediction of compressive strength of concrete containing rice husk ash, the optimal number of individuals per generation was set to 30 chromosomes, and the best fitness value achieved was 633.68. Fig. 4 presents the prediction deviations of the GEP model in comparison to the experimentally obtained CS values of self-compacting concrete. A higher model accuracy is indicated by data points clustering near the zero-error line. As shown, most of the errors lie within the interval of -8 to +8 MPa.

Table 5. Optimal configuration parameters for the GEP model.

Parameter details		Parameter details	
Function set	+, −, /, exp, power	Gene transposition rate	0.277
Mutation rate	0.044	Head size	7
Inversion rate	0.3	Number of genes	4
One/two-point recombination rate	0.3	Number of chromosomes	30
Gene recombination rate	0.1		

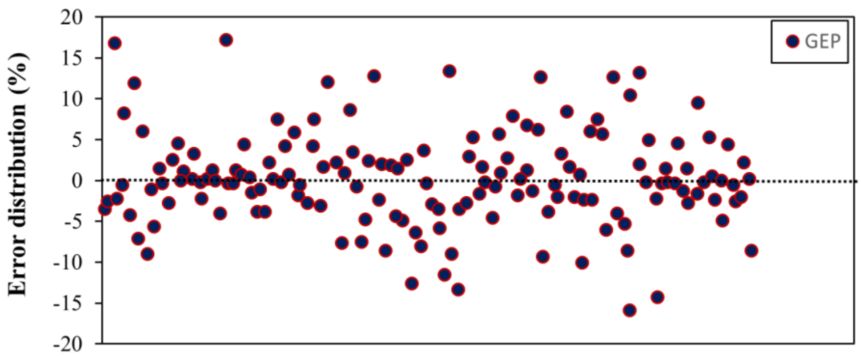
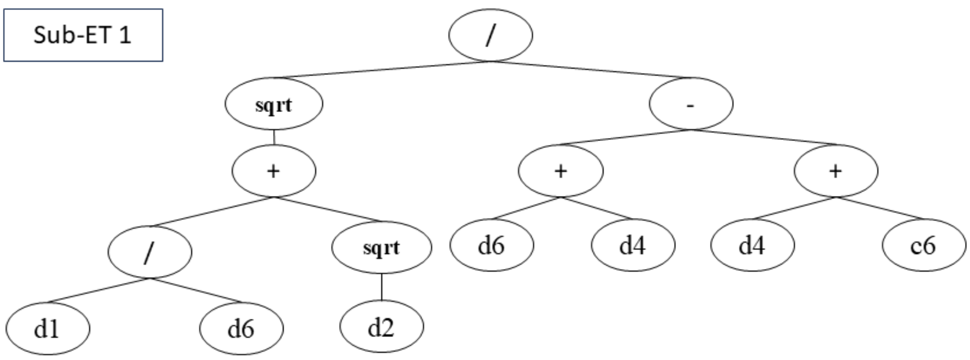


Fig. 4. Distribution of prediction errors in the GEP model.

Based on previous studies and through a process of trial and error, the optimal combination of input parameters for predicting the compressive strength of concrete containing rice husk ash was identified. According to Table 5 and the structure shown in Fig. 5, the GEP-based model was developed using a selection of parameters that significantly influence the target output. The notations used are: d0 = C, d1 = RHA, d2 = W, d3 = SP, d4 = FA, d5 = CA, and d6 = AS. The final input configuration selected C, W, RHA, SP, FA, CA, and AS, indicating that rice husk ash content, water-to-cement ratio, and superplasticizer dosage are the most influential variables, along with the type and size of aggregates and curing age.



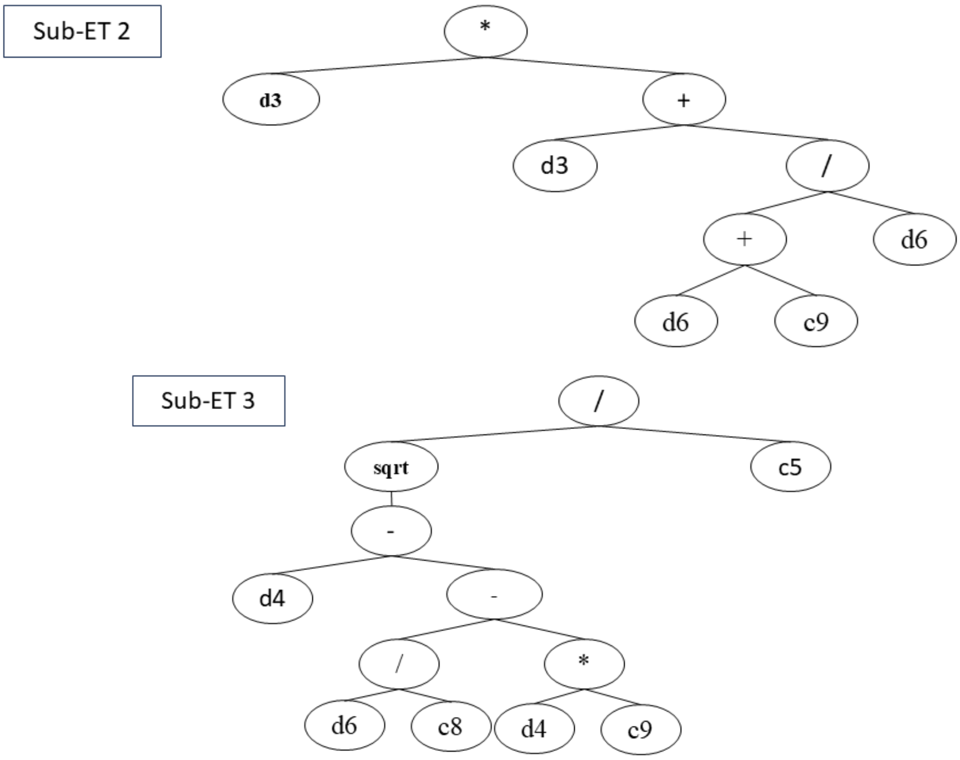


Fig. 5. Expression tree of GEP model.

Finally, the best predictive equation for estimating the CS of SCC with RHA, based on the optimal combination of input parameters, is expressed through Eq. (1).

$$\mathrm{CS} = \left\{\left(\left(\frac{RHA}{C}\right) + W^{0.5}\right)^{0.5} \times \left((AS + FA) - \ (FA + 3.49)^{-1}\right)\right\} + \left\{SP \times \left(CA + (AS - 7.26)\right)\right\} + \left\{\left(FA + \left(\frac{AS}{0.234}\right) - 7.9AS\right)^{0.5} + 6.8\right\} \tag{1}$$

As shown in Table 6, the R in the training phase for MLR, ANN, and GEP models was 0.92, 0.95, and 0.98, respectively. The RMSE values were 6.4, 4.5, and 3.9, and the MAE values were 7.8, 5.1, and 5.4, respectively. These results indicate that the GEP model outperformed ANN and MLR in terms of prediction accuracy.

The ranking of models based on error metrics shows that GEP had the highest accuracy and required less training time compared to ANN. This suggests that GEP is a more efficient and reliable approach, particularly when interpretability of the prediction model is important. Considering the mathematical structure of the GEP-derived equation and the strong consistency between predicted and observed values, it can be inferred that the GEP model delivers highly precise outcomes, with prediction errors remaining within acceptable limits.

Table 6. Comparative analysis of developed models for predicting the CS of SCC.

Models	Training			Testing		
	R	RMSE	MAE	R	RMSE	MAE
MLR	0.92	6.4	7.8	0.9	8.2	9.5
ANN	0.95	4.5	5.1	0.93	5.1	5.6
GEP	0.98	3.9	5.4	0.95	4.3	4.9
Belalia Douma et al. [28]	0.95	-	-	0.94	-	-
Kaveh et al. [29]- a	0.95	5.2	5.1	0.95	6.8	5.5
Kaveh et al. [29] - b	0.96	4.5	3.5	0.94	4.5	5.6
Inqiad et al. [30]	0.22	733	23.1	0.33	4.5	5.6
Siddique et al. [31]	-	-	-	0.91	4.43	5.6

In this study, Fig. 6 illustrates the observed and predicted compressive strength values, as estimated by the proposed MLR, ANN, and GEP models during the training phase. A comparative qualitative assessment shows that the GEP method more accurately tracks the observed compressive strength values. In contrast, the MLR model tends to diverge from the data points, especially in regions with high error. The GEP method exhibits greater consistency and lower dispersion around the regression line, focusing on a denser clustering of points near the actual values. The MLR model, on the other hand, displays a wider spread in the error range, indicating reduced accuracy in certain regions. Fig. 7 further confirms that the GEP model yields predictions that are more closely aligned with experimental results. Additionally, it was observed that during the testing phase, the GEP model achieved lower error values compared to ANN, highlighting the superior accuracy of gene expression programming over artificial neural networks in this case.

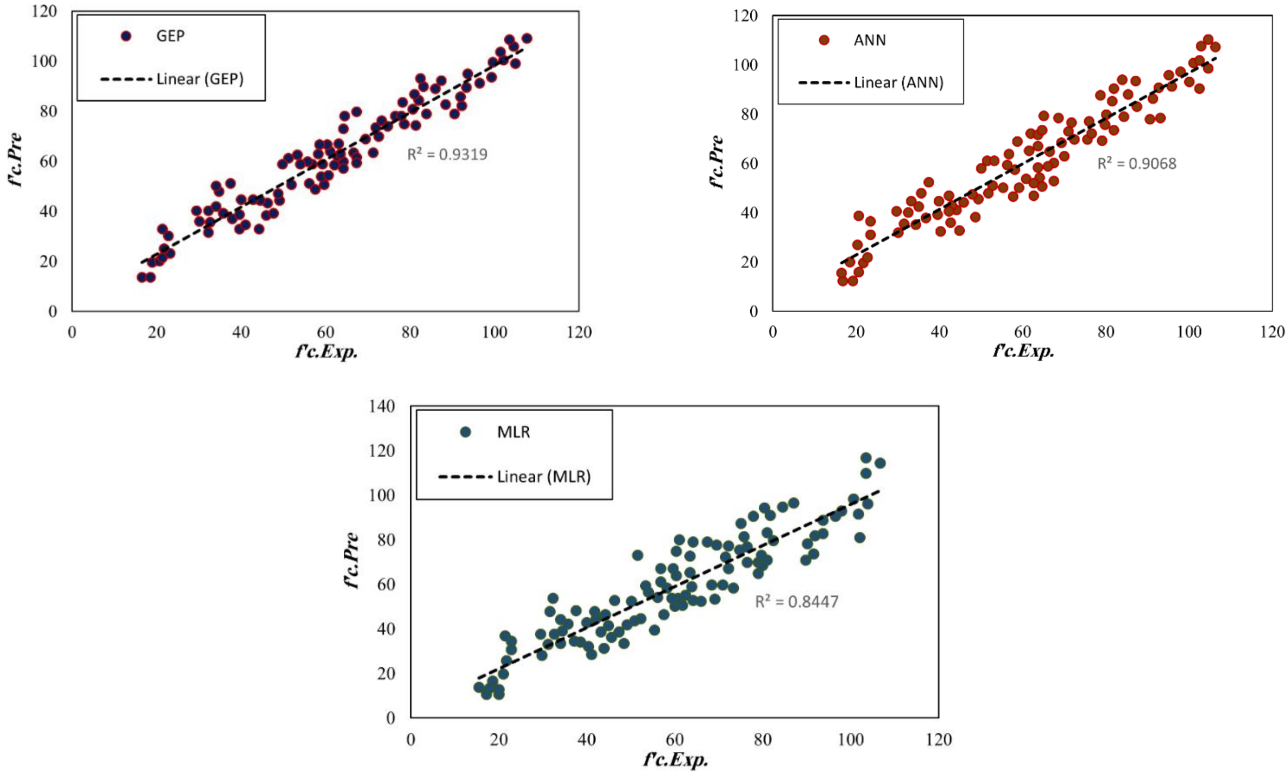


Fig. 6. Scatterplots of the developed models for training performance.

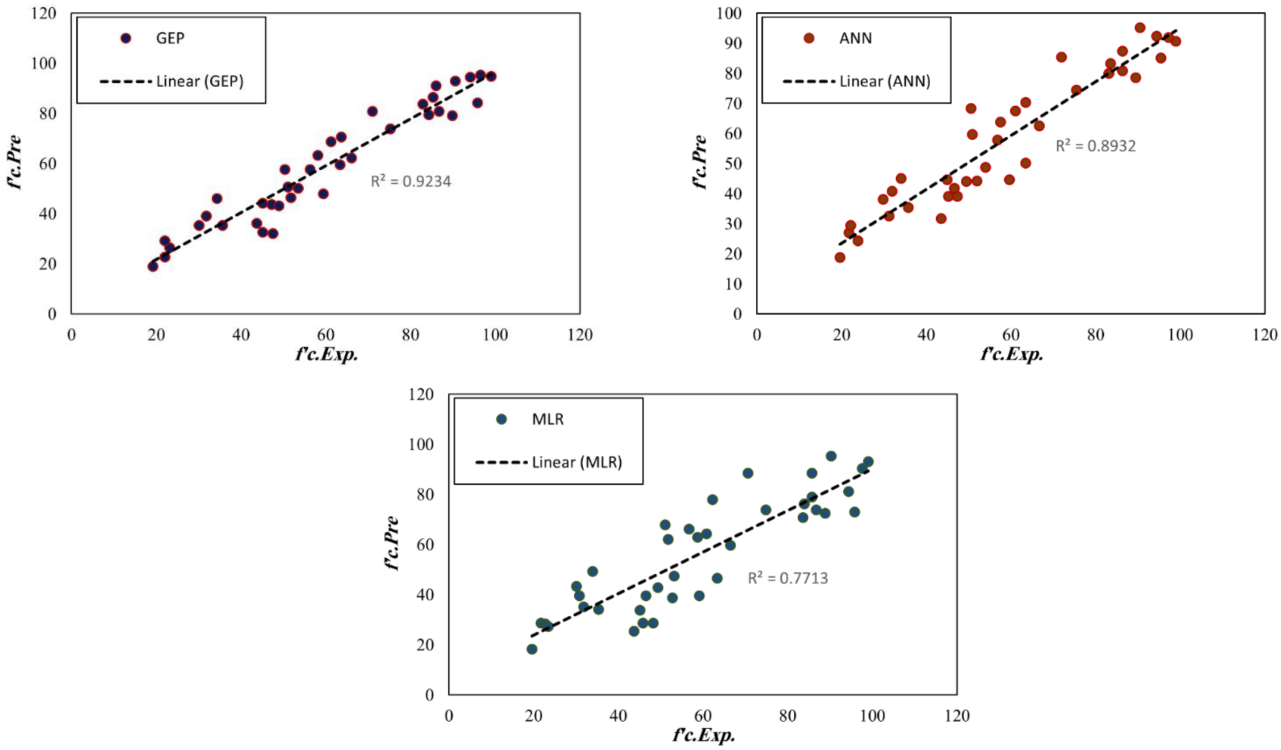


Fig. 7. Scatterplots of the developed models for testing performance.

Traditional approaches like ANNs and MLR were employed in this research to estimate the CS of SCC incorporating RHA. From a comparative standpoint, the quantitative evaluation results shown in Fig. 8 indicate that the proposed methods outperform traditional approaches. The GEP method, in particular, exhibited greater accuracy and lower error dispersion than the MLR model. Because linear regression cannot model intricate nonlinear patterns, it failed to reliably predict and generalize compressive strength values over the full spectrum of the dataset. On the other hand, the GEP-based model provided more accurate predictions and minimized the deviation from experimental results, especially when determining the influence of multiple variables. Additionally, previous studies [23] have emphasized that methods such as ANN are highly sensitive to the number of variables and input conditions, which may lead to instability in model behavior. Therefore, based on the results of Fig. 8, it can be concluded that the GEP method offers higher predictive performance for compressive strength estimation compared to both ANN and MLR.

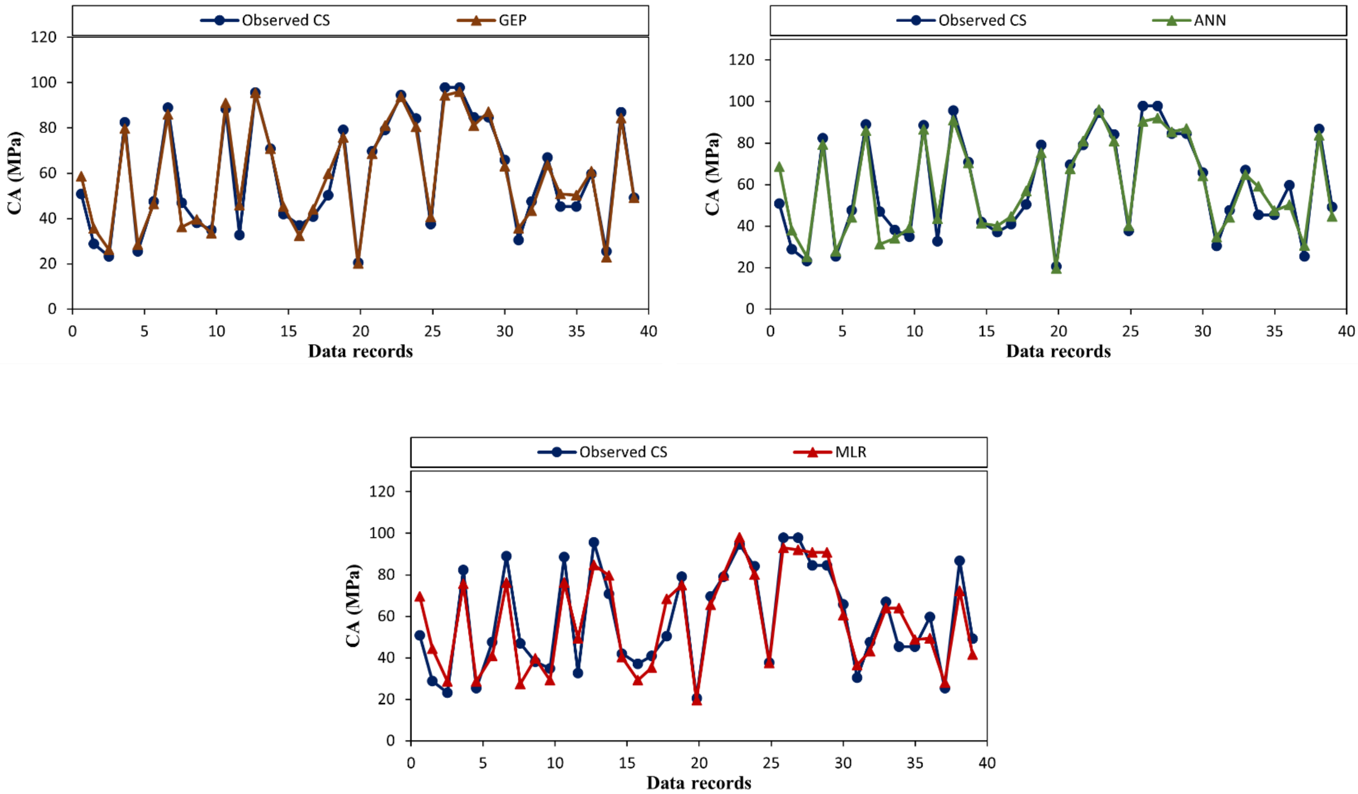


Fig. 8. Time series plots of model performance in the testing phase.

In this study, classical data-driven methods such as ANN and MLR encountered computational errors due to the trial-and-error nature of network optimization and their limited ability to address complex problems involving a large number of variables affecting the physical behavior of different systems [23]. Therefore, the use of metaheuristic algorithms has become essential for improving these methods and other predictive approaches. According to the timeline chart illustrated in Fig. 8, the CS predictions made using the MLR method demonstrated relatively weaker performance compared to the ANN and GEP models in predicting the CS of SC with RHA. The quantitative comparison shown in Fig. 8 confirms that the MLR method produced relatively inaccurate predictions and failed to effectively estimate both the maximum and minimum local values of compressive strength. The highest prediction errors were observed in the 40–60 MPa range, where the predicted values were generally lower than the experimentally measured compressive strengths. This highlights the MLR model's inability to capture the nonlinear relationships involved in concrete strength development when using supplementary cementitious materials like rice husk ash.

5.4. Assessment of the relevance of the developed regression models

To test the significance of the regression coefficients, statistical testing using the F-value can be employed as follows:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ → the regression model is not significant.

H_1: $\beta_i \neq 0 : i = 1, 2, \ldots, k$ → the regression model is significant.

If all the regression coefficients are equal to zero, this would indicate that there is no linear relationship between the independent variables and the dependent variable. In that case, the model would be deemed statistically insignificant. If the result of the test indicates that at least one coefficient differs from zero, the regression model is considered significant. To test the significance of the regression model, the confidence level must be determined first. In this study, a 95% confidence level was selected, which means the probability of rejecting the null hypothesis should be less than 0.05. The critical value for model significance evaluation using the Fisher test [24] is determined based on Table 7. It should be noted that the F-test results reported in Table 8 were obtained based on the training dataset. The F-statistic and p-values were calculated in Python using the statsmodels library, which provides an automated regression analysis. This library computes the regression sum of squares, residual sum of squares, mean squares, and finally the F-value, following standard statistical procedures. In the results of model significance testing, the p-value is reported. If the p-value is less than the significance level (0.05), the null hypothesis is rejected, and the model is considered statistically significant. Also, as shown in the Table of statistical analysis results, the reported p-values are all less than 0.05, which confirms the significance of the regression model. Furthermore, the F-value reported is higher than the critical value (9.16), indicating a strong model fit.

Table 7. F-test outcomes for the developed models.

Model	F-value	P-value
GEP	22.11	0.0088
ANN	15.1	0.02
MLR	8.4	0.38

5.5. Sensitivity analysis

Sensitivity analysis refers to the study of how the input variables of a statistical model influence its output. In other words, it is a systematic method of varying the inputs of a model to predict the effects of these changes on the model's output [25]. In this research, sensitivity analysis was conducted using the GEP approach, which showed the highest accuracy among all developed models to evaluate the impact of each input variable on the predicted CS of SCC incorporating RHA.

The results presented in Table 8 indicate that the parameter cement (C), with a correlation coefficient of R = 0.73 and other statistical indicators (RMSE = 10.4, MAE = 11.3), has the greatest influence on predicting the compressive strength of concrete. The parameter aggregate size (AS) also shows a high correlation with compressive strength, with a correlation coefficient of R = 0.93, and statistical indicators (RMSE = 4.2, MAE = 4.2) demonstrating its strong predictive influence.

Table 8. Sensitivity analysis of input variables for predicting CS.

Imput	R	RMSE	MAE
CS = *f*(W, SP, FA, CA, AS, RHA)	0.73	10.4	11.3
CS = f(C, SP, FA, CA, AS, RHA)	0.89	5.3	6.1
CS = f(C, W, FA, CA, AS, RHA)	0.92	4.2	4.2
CS = f(C, W, SP, CA, AS, RHA)	0.9	4.7	5.3
CS = f(C, W, SP, FA, AS, RHA)	0.79	9.7	10.4
CS = f(C, W, SP, FA, CA, RHA)	0.93	4.2	4.2
CS = f(C, W, SP, FA, CA, AS)	0.82	7.7	7.8

These findings confirm that cement and aggregate size are the two most influential parameters in determining the CS of SCC incorporating RHA [32-34]. In contrast, other input variables W, RHA, CA, FA, and SP had comparatively less influence. Overall, the sensitivity analysis revealed that cement and aggregate size are the dominant factors affecting the compressive strength of SCC with RHA.

6. Conclusion

In this study, one of the most advanced data-driven techniques, GEP, was employed to predict the CS of SCC incorporating RHA. To identify the most influential parameters affecting the compressive strength of this concrete type, four input scenarios were defined, and each case was evaluated using MLR. Ultimately, seven input parameters (C, AS, CA, FA, SP, W, and RHA) and one output parameter (concrete compressive strength) were used to develop the proposed models.

Based on the evaluation of the developed models, the following conclusions were drawn:

1. The performance of the proposed intelligent models was assessed for both training and testing phases using experimental data collected from previous studies. According to the statistical error indicators in the training phase, the GEP model (R = 0.97, RMSE = 3.364) outperformed the Artificial Neural Network model (R = 0.94, RMSE = 4.536) and the Multiple Linear Regression model (R = 0.91, RMSE = 6.762) in predicting the compressive strength of SCC. Furthermore, performance evaluation in the testing phase showed that the GEP model (R = 0.97, RMSE = 3.308) provided more accurate predictions under nonlinear formulations compared to the ANN (R = 0.92, RMSE = 5.136) and MLR (R = 0.89, RMSE = 9.212) models for the 21-day compressive strength of SCC containing RHA.

2. The GEP and ANN models, used as explicit equations for estimating the compressive strength of SCC, demonstrated significantly higher accuracy compared to traditional linear methods. In particular, GEP, with its capability to formulate precise output estimations, proved to be a powerful tool for prediction and data mining in engineering technologies, especially in concrete technology.

3. The statistical significance of the proposed models was verified using regression hypothesis testing through the Fisher test, showing that the F-values for all models were below the critical threshold and the confidence levels were less than 0.05, thereby validating the reliability of the developed models.

4. To assess the sensitivity of the output variable to the input variables, the GEP method was used. The results indicated that variable C had the most substantial effect on predicting compressive strength, whereas SP exhibited the least influence on the model’s output.

5. Based on the results of the sensitivity analysis and the performance of the developed models, the optimal RHA content for

achieving higher compressive strength in self-compacting concrete mixtures was found to be approximately 10–15% by weight of the binder.

Statements & Declarations

Author contributions

Hamid Farrokh Ghatte: Conceptualization, Validation, Visualization, Writing - review & editing.

Ali Nazari: Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Writing - Original Draft.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Data availability

The data presented in this study will be available on interested request from the corresponding author.

Declarations

The authors declare no conflict of interes.

References

[1] Sandhu, R. K., Siddique, R. Influence of rice husk ash (RHA) on the properties of self-compacting concrete: A review. Construction and Building Materials, 2017; 153: 751-764. doi:10.1016/j.conbuildmat.2017.07.165.

[2] Ameri, F., Shoaei, P., Bahrami, N., Vaezi, M., Ozbakkaloglu, T. Optimum rice husk ash content and bacterial concentration in self-compacting concrete. Construction and Building Materials, 2019; 222: 796-813. doi:10.1016/j.conbuildmat.2019.06.190.

[3] Sutas, J., Mana, A., Pitak, L. Effect of Rice Husk and Rice Husk Ash to Properties of Bricks. Procedia Engineering, 2012; 32: 1061-1067. doi:10.1016/j.proeng.2012.02.055.

[4] Qu, J., Li, B., Wei, T., Li, C., Liu, B. Effects of rice-husk ash on soil consistency and compactibility. CATENA, 2014; 122: 54-60. doi:10.1016/j.catena.2014.05.016.

[5] Chen, L., Pai, T. Y. An ERAFNN Prediction of Concrete Compressive Strength Based on Physical Properties of Electric Arc Furnace Oxidizing Slag. Applied Mechanics and Materials, 2014; 627: 111-114. doi:10.4028/www.scientific.net/AMM.627.111.

[6] Sonebi, M., Grünewald, S., Cevik, A., Walraven, J. Modelling fresh properties of self-compacting concrete using neural network technique. Computers and Concrete, 2016; 18: 903-921. doi:10.12989/cac.2016.18.4.903.

[7] Kalman Šipoš, T., Miličević, I., Siddique, R. Model for mix design of brick aggregate concrete based on neural network modelling. Construction and Building Materials, 2017; 148: 757-769. doi:10.1016/j.conbuildmat.2017.05.111.

[8] Wang, B., Man, T., Jin, H. Prediction of expansion behavior of self-stressing concrete by artificial neural networks and fuzzy inference systems. Construction and Building Materials, 2015; 84: 184-191. doi:10.1016/j.conbuildmat.2015.03.059.

[9] Rezaei, F., Memarzadeh, A., Davoodi, M.-R., Dashab, M.-A., Nematzadeh, M. Mechanical features and durability of concrete incorporating recycled coarse aggregate and nano-silica: Experimental study, prediction, and optimization. Journal of Building Engineering, 2023; 73: 106715. doi:10.1016/j.jobe.2023.106715.

[10] Gholampour, A., Gandomi, A. H., Ozbakkaloglu, T. New formulations for mechanical properties of recycled aggregate concrete using gene expression programming. Construction and Building Materials, 2017; 130: 122-145. doi:10.1016/j.conbuildmat.2016.10.114.

[11] Nematzadeh, M., Nazari, A., Tayebi, M. Post-fire impact behavior and durability of steel fiber-reinforced concrete containing blended cement–zeolite and recycled nylon granules as partial aggregate replacement. Archives of Civil and Mechanical Engineering, 2021; 22: 5. doi:10.1007/s43452-021-00324-1.

[12] Nematzadeh, M., Tayebi, M., Samadvand, H. Prediction of ultrasonic pulse velocity in steel fiber-reinforced concrete containing nylon granule and natural zeolite after exposure to elevated temperatures. Construction and Building Materials, 2021; 273: 121958. doi:10.1016/j.conbuildmat.2020.121958.

[13] Sobhani, J., Najimi, M., Pourkhorshidi, A. R., Parhizkar, T. Prediction of the compressive strength of no-slump concrete: A comparative study of regression, neural network and ANFIS models. Construction and Building Materials, 2010; 24: 709-718. doi:10.1016/j.conbuildmat.2009.10.037.

[14] Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 1958; 65: 386. doi:10.1037/h0042519.

[15] Augustino, D. S. Parametric Modelling of Fibre-Concrete Interface and Prediction of Compressive Strength of Concrete With Waste Tyre Steel Fibres Using Artificial Neural Network Model. Advances in Civil Engineering, 2025; 2025: 7678622. doi:10.1155/adce/7678622.

[16] Marimuthu, V., Ramasamy, A. Mechanical characteristics of waste-printed circuit board-reinforced concrete with silica fume and prediction modelling using ANN. Environmental Science and Pollution Research, 2024; 31: 28474-28493. doi:10.1007/s11356-024-33099-y.

[17] Ferreira, C. Gene expression programming: a new adaptive algorithm for solving problems. arXiv preprint cs/0102027, 2001; doi:10.48550/arXiv.cs/0102027.

[18] Shahmansouri, A. A., Akbarzadeh Bengar, H., Jahani, E. Predicting compressive strength and electrical resistivity of eco-friendly concrete containing natural zeolite via GEP algorithm. Construction and Building Materials, 2019; 229: 116883. doi:10.1016/j.conbuildmat.2019.116883.

[19] Sakr, K. Effects of silica fume and rice husk ash on the properties of heavy weight concrete. Journal of materials in civil engineering, 2006; 18: 367-376. doi:10.1061/(ASCE)0899-1561(2006)18:3(367).

[20] Ganesan, K., Rajagopal, K., Thangavel, K. Rice husk ash blended cement: Assessment of optimal level of replacement for strength and permeability properties of concrete. Construction and Building Materials, 2008; 22: 1675-1683. doi:10.1016/j.conbuildmat.2007.06.011.

[21] Bui, D. D., Hu, J., Stroeven, P. Particle size effect on the strength of rice husk ash blended gap-graded Portland cement concrete. Cement and Concrete Composites, 2005; 27: 357-366. doi:10.1016/j.cemconcomp.2004.05.002.

[22] Rodríguez de Sensale, G. Strength development of concrete with rice-husk ash. Cement and Concrete Composites, 2006; 28: 158-160. doi:10.1016/j.cemconcomp.2005.09.005.

[23] Shahmansouri, A. A., Yazdani, M., Hosseini, M., Akbarzadeh Bengar, H., Farrokh Ghatte, H. The prediction analysis of compressive strength and electrical resistivity of environmentally friendly concrete incorporating natural zeolite using artificial neural network. Construction and Building Materials, 2022; 317: 125876. doi:10.1016/j.conbuildmat.2021.125876.

[24] Shahmansouri, A. A., Yazdani, M., Ghanbari, S., Akbarzadeh Bengar, H., Jafari, A., Farrokh Ghatte, H. Artificial neural network model to predict the compressive strength of eco-friendly geopolymer concrete incorporating silica fume and natural zeolite. Journal of Cleaner Production, 2021; 279: 123697. doi:10.1016/j.jclepro.2020.123697.

[25] Ghatte, H. F. Evaluation of reinforcing bars ratio effects on SCC beam-column joint performance. Uludağ Üniversitesi Mühendislik Fakültesi Dergisi, 2019; 24: 141-152. doi:10.17482/uumfd.587505.

[26] Farrokh Ghatte, H. External steel ties and CFRP jacketing effects on seismic performance and failure mechanisms of substandard rectangular RC columns. Composite Structures, 2020; 248: 112542. doi:10.1016/j.compstruct.2020.112542.

[27] Farrokh Ghatte, H. Failure mechanisms and cracking performance of T-shaped SCC beam-column connections at top floor: Test results and FE modeling. Structures, 2020; 28: 1009-1018. doi:10.1016/j.istruc.2020.09.051.

[28] Belalia Douma, O., Boukhatem, B., Ghrici, M., Tagnit-Hamou, A. Prediction of properties of self-compacting concrete containing fly ash using artificial neural network. Neural Computing and Applications, 2017; 28: 707-718. doi:10.1007/s00521-016-2368-7.

[29] Kaveh, A., Bakhshpoori, T., Hamze-Ziabari, S. M. M5'and Mars based prediction models for properties of self-compacting concrete containing fly ash. Periodica Polytechnica Civil Engineering, 2018; 62: 281-294. doi:10.3311/PPci.10799.

[30] Inqiad, W. B., Siddique, M. S., Ali, M., Najeh, T. Predicting 28-day compressive strength of fibre-reinforced self-compacting concrete (FR-SCC) using MEP and GEP. Scientific Reports, 2024; 14: 17293. doi:10.1038/s41598-024-65905-5.

[31] Siddique, R., Aggarwal, P., Aggarwal, Y. Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks. Advances in Engineering Software, 2011; 42: 780-786. doi:10.1016/j.advengsoft.2011.05.016.

[32] Zhang, W. G., Goh, A. T. C. Multivariate adaptive regression splines for analysis of geotechnical engineering systems. Computers and Geotechnics, 2013; 48: 82-95. doi:10.1016/j.compgeo.2012.09.016.

[33] Bluman, A. G. Elementary statistics: a step by step approach. 11th ed. New York (NY): McGram-Hill; 2019.

[34] Najafzadeh, M., Rezaie Balf, M., Rashedi, E. Prediction of maximum scour depth around piers with debris accumulation using EPR, MT, and GEP models. Journal of Hydroinformatics, 2016; 18: 867-884. doi:10.2166/hydro.2016.212.